INTERNATIONAL JOURNAL OF COMPUTATIONAL MATHEMATICAL IDEAS ISSN: 0974-8652

Tracking Temporal Community Strength In Dynamic Networks

Mr.1 Polisetty A Satish Kumar Mr.2 G CH Srinivasa Rao
Newton Institute of Engineering

Published in Volume 13, Issue 1 Oct - Nov: 2016, Page No: 65050 to 65060

Abstract:

Community formation analysis of dynamic networks has been a hot topic in data mining which has attracted much attention.Recently, there are many studies which focus on discovering communities successively from consecutive snapshots by considering both the current and historical information. However, these methods cannot provide us with much historical or successive information related to the detected communities. Different from previous studies which focus on community detection in dynamic networks, we define anew problem of tracking the progression of the community strength - a novel measure that reflects the community robustness and coherence throughout the entire observation period. To achieve this goal, we propose a novel framework which formulates the problem as an optimization task. The proposed community strength analysis also provides foundation for a wide variety of related applications such as discovering how the strength of each detected community changes over the entire observation period. To demonstrate that the proposed method provides precise and meaningful evolutionary patterns of communities which are not directly obtainable from traditional methods, we perform extensive experimental studies on one synthetic and five real datasets: social evolution, tweeting interaction, actor relationships, bibliography and biological datasets. Experimental results show that the proposed approach is highly effective in discovering the progression of community strengths and detecting interesting community

KEYWORDS:

Dynamic Networks, Community Analysis, Community Strength.

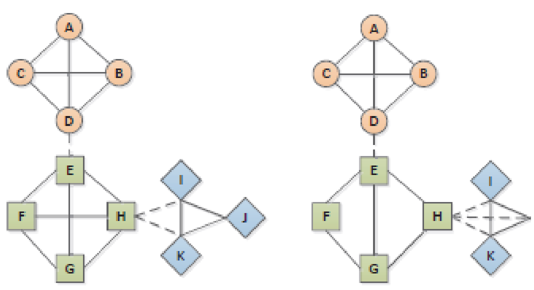
INTRODUCTION:

In recent years, there has been a growing interest in modeling and mining various kinds of dynamic networks such as biological networks, social networks, co-author ship networks and co-starring networks whose structures evolve over time. Among extensive work, community analysis in dynamic networks which focuses on detecting the communities successively from each snapshot by considering the historical information has recently attracted much attention However, all these detected communities are frozen and isolated at a specific snapshot. Thus we do not know when these communities were assembled or when they are going to disband. Aiming to answer these questions, we propose a novel measure called *community strength*. In this paper, we claim that a community is with high strength if it has more internal interactions connecting the members of the community than the external interactions connecting to the rest of the world. Dense internal interactions and weak external interactions to the outside guarantee that the community is under a low risk of member change (current members leaving or/and new members joining).It is easy to understand that a friend community is somewhat strong if its members tie together closely and ignore the temptation from the outside world. On the contrary, we regard a friend community as a weak community, if it is likely to confront a member alteration situation. To illustrate this idea, Fig*:* 1(a) shows a toy example, where the nodes represented by the same geometric shape belong to the same community, solid lines represent internal interactions and dash lines represent external interactions. The circle community (i.e. nodes A, B, C and D) is considered to be stronger than the rectangle community (i.e. nodes E, F, G and H), because it involves less external attractions. Specifically, node H is having a close relationship with the diamond community (i.e. nodes I, J and K), which makes the rectangle community in the risk of losing its members. The higher strength score a community obtains, the less possible member alternation occurs in it

RELATED WORK:

Methodology

The objective of this study is to track the evolution of communities over time in dynamic social networks. We represent a social network by an undirected weighted graph, where the nodes of the graph represent the members of the network, and the edge weights represent the strengths of social ties between members. The edge weights could be obtained by observations of direct inter- action between nodes, such as physical proximity, or inferred by similarities between behavior patterns of nodes. We represent a dynamic social network by a sequence of time snapshots, where the snapshot at time step *t* is represented by *Wt* =[*wtij*], the matrix of edge weights at time *t*. *Wt* is commonly referredto as the adjacency matrix of the network snapshot.The problem of detecting communities in static networks has been studied by researchers from a wide range of disciplines. Many community detection methods originated from methods of graph partitioning and data clustering. Popular community detection methods include modularity maximization and spectral clustering [12,14]. In this paper, we address the extension of community detection to dynamic networks, which we call *c*



(a) 1^{st} *snapshot* (b) 2^{nd} *snapshot*

ommunity tracking.

Discovering the progression of community strengths can offer significant insights in a variety of applications. It can help us discover some interesting community information which can not be directly obtained from traditional community analysis. Interesting examples of communities' strength progression can be commonly observed in real-life scenarios. Here we discuss two specific cases in detail.

Strengths Progression in Actor Community:

As a strong actor community, the cooperation should be more frequent between the members themselves than between members and non-members. For example, considering the popular and long-running television sitcom Friends'1, its six main actors J. Aniston, C. Cox, M. Perry,M. LeBlanc, L. Kudrow and D. Schwimmer collaborated closely when this sitcom was aired from 1994 to 2004. Let's consider each year's co-starring relationships as one snapshot. We can see that the strength of this community is very low before 1994 (little cooperation between them), and then dramatically increases and keeps stable from 1994 to 2004 (average 23 episodes each year). Finally, the strength of this community apparently becomes weaker after 2004 (much less cooperation comparing to the previous years). The progression of this actor community's strengths shows an interesting pattern of cooperation history among these six actors. Learning the strength progression of actor communities helps us better understand the entertainment industry.

Strength Progression in Gene Community:

In the biological domain, the interactions between genes change gradually in dynamic gene co-expression networks. Thus the strength of gene communities also changes. For example, it has been reported that the expression profiling of some key genes will change [8] as the cancer progresses. In such cases, the corresponding gene communities' strength also changes. Discovering the strengths of gene communities throughout a specific disease progression can help us find significant clues in the fields of medicine and biology. For a specific disease, if a gene community is found strong only at the early stage, it is very likely to be a crucial trigger for the disease deterioration. From the above cases, we can see that discovering the progression of community strengths helps us understand the underlying behavior of communities. The initial idea was published in [9], which covers the basic definition of community strength and the evolutionary analysis on dynamic networks. By utilizing the community strength value, the consistent communities can be detected and tracked over an observation period. This paper extends the original idea to formulate a solid method with broader applications and provide more supportive and comprehensive experiments. In this paper, our goal is to detect the temporal strength of each detected community throughout all the snapshots so that we can answer the following questions: How does the strength of each community change over the observation period? What are the top-K strong communities throughout the observation period? How do the communities from adjacent snapshots influence the strength of each other?

To sum up, our main contributions in this paper are as follows:

We introduce the notion of progression analysis of community strengths. To the best of our knowledge, this is the first work on analyzing the temporal community quality or structure information considering both time and community information.We formulate the problem as an optimization framework that can effectively detect the temporal strength of communities and track the strength progression pattern. Experiments on the synthetic dataset show the proposed approach is effective on identifying strong communities. On real datasets, interesting and meaningful communities are detected. Case studies suggest that the proposed approach can provide more reasonable results.The organization of the paper is as follows: In Section2, we describe the setting of our problem. Section 3presents the analysis and discussions related to the proposed algorithm.at snapshot *t*.

METHODOLOGY:

In this section, we present our method for solving the problem of temporal community strength analysis. We begin by introducing the method of partitioning the network from each snapshot into communities in Section III-A, and show the methods of tracking the strength of each community at each snapshot and post processing work of our framework to handle some practical issues in the a series of temporal networks *Gt* = (*V t;Et;Wt*) (1 ≤ *t* ≤ *T*), we first partition each network independently into multiple communities, and then collect all the detected communities from every snapshot together as a community pool. To detect communities at each temporal network, we use the Non-negative Matrix Factorization (NMF) [5], which is widely used in many applications such as pattern recognition, signal modeling, bioinformatics and text mining. There are two reasons to use NMF: first, it can be easily applied to both hard clustering (i.e. each object belongs to exactly one community) and soft clustering (i.e. each object can belong to multiple communities); second, it could uncover more precise underlying similarities between communities. Given a network *Wt* ∈ ℜ*N1N*, it can be decomposed into two components: *Ct* ∈ ℜ*N1K* and *St* ∈ ℜ*K1K* as:min *Ct*0*;St*0 *Wt* − *CtStCtT* 2 *; s:t: CtT Ct* = *I;* (2) where *Wt* is an *N* × *N* symmetric matrix that demonstrates the interactions between objects at time *t*, *Ct* is an *N* × *Kt* community indicator matrix representing the probability of grouping an object into a community and *St* is a *Kt* × *Kt* non-negative matrix providing the relationship between com- munities detected at time *t*. The basic idea of minimizing Eq*:* 2 is that, at iteration *t* we first fix *St* and update *Ct* as: *ctij*←*ctij*2√(*WtT CtStT*)*ij*(*CtCtTWTCtStT*)*ij:* (3)Similarly, fixing *Ct*, we can obtain the updated rule for *St* as:*stij*← *stij*2√*CtTWtT Ct*)*ij*(*CtT CtStCtT Ct*)*ij:* (4)We iteratively update *Ct* and *St* until convergence. It is worth noticing that the community indicator matrix *Ct* can be applied to both hard clustering and soft clustering. Since it is hard to tell whether a community is strong or not without comparing with other communities, we put all the unique detected communities together into a community pool ~ *CN1K*. This community pool is used as a candidate set where all the communities are easily normalized and compared. Based on it, we intend to find out which communities are grouped closely and consistently and which communities are grouped temporarily. Note that, although we remove the duplicate communities which are exact the same in our case, other filtering methods could be used. For example, a threshold can be set that, if any pair of communities have a similarity over this threshold, we merge these two communities.

B. Temporal Community Strength Analysis

Now, we propose an integrated optimization framework that conducts community strength detection across snapshots. There are mainly two reasons why we use an integrated optimization framework rather than calculate the strength of each community individually at each snapshot. First, our framework is based on the smoothness assumption in which both current and historical network contribute to the community strength detection.

Community Detection at Each Snapshot:

Given a series of temporal networks Gt = (V;Et;Wt) (1 t T), we first partition each network independently into Kt communities at each timestamp t. Due to the change of network, the value of Kt may not be the same across different snapshots.Then we store all the detected communities from all the snapshots in a community pool. To detect communities from each temporal network, we use Non-negative Matrix Factorization (NMF) technique There are two major reasons to choose NMF: First, it can be easily applied to both hard clustering (i.e. each object belongs to exactly one community) and soft clustering (i.e. each object can belong to multiple communities). The property of soft clustering very well fits many real social scenarios. For instance, each user in social network usually participates in more than one discussion group, as he may have a variety of interested.

Temporal Community Strength Analysis

Now, we propose an integrated optimization framework that conducts community strength estimation across snapshots. A naive approach for this task is to calculate the strength of each community individually at each snapshot and track the evolution. However, this approach does not take historical information into account when deriving community strengths and the communities derived across snapshots are not easily comparable. In contrast, we propose the following framework based on the smoothness assumption in which both current and historical networks contribute to the community strength detection. Moreover, in the proposed framework, communities across snapshots are brought into alignment so that we can easily compare them. Based on Eq. 1, the strength of community z can be further reformulated in terms of the community pool matrix ~ C as follows:

Community Strength Progression Net:

The output of Algorithm 1 provides information on how all the communities' strength evolve over time. In addition to that, we also want to know how the communities from immediate preceding snapshots (i.e. Ct□1 and Ct) influence the strength of each other. To illustrate these relationships, we construct a bipartite network that represents the relationship between communities detected at snapshot t-1 and communities detected at snapshot t. In such a network, the nodes on the left represent the communities detected at previous time stamp, the nodes on the right represent the communities detected at the current time stamp and the edges connecting the nodes denote the influence transmission between the communities. The relationship matrix PKt□1 1 Kt that represents the relationships between communities captured at adjacent snapshots (t-1 and t) can be calculated as: P = D□1St□1Ct□1CtT StT ; (11) where D is a diagonal matrix used for normalization and Dii = PKt j=1(St□1Ct□1CtT StT)ij . As we mentioned in Section 3.1, Ct and Ct□1 are the community indicator matrices with respect to snapshot t and t-1. St and St□1 represent the relationship between communities at snapshot t and t-1, respectively. As we mentioned previously, St can uncover the underlying relationships between communities detected at snapshot t, and Ct□1CtT demonstrates the number of common members between the two snapshots' communities. Thus, P can reflect not only the common member relationship but also the underlying relationships between two snapshots' communities. A natural definition of community progression net (from ct□1 i at time t-1 to ct j at time t) is a flow starting from ct□1 i , and transmits its strength to ct j . There are two applications that are worth discussing: First, we analyze how the community strength from the current snapshot transmits to the next snapshot. Second, we analyze how the current community strength succeeds from the previous snapshot. For the former one, the strength transmits community i at the current snapshot to the community j at the next snapshot, which is defined as a it pij . As mentioned before, ait is the strength of community I at time t and pij is the relationship among community i and j, aitpij can reflect the influence community j obtained from community i. The network reflecting this transmission relationship is named Strength Transmission Net. Correspondingly, for the latter one, the strength that the current community j inherits from community i is defined as pijajt, which is named Strength Reception Net. Notice that to measure the Strength Reception Net, we need to normalize each column of P.Examples for Strength Transmission Net and Strength Reception Net are depicted in Fig. 2(a) and 2(b), respectively.

In each network, the values shown inside the geometric shapes are the strength corresponding to the communities. For example, from Fig. 2(a) we can see that the circle community from the 1st snapshot transmits its current strength (0.46) to the succeeding circle community with 0.44 and rectangle community with 0.02. Take another example, from Fig. 2(b), we can see the diamond community at the 2nd snapshot inherits 0.35 and 0.03 strength from diamond community and rectangle community at the 1st snapshot. In such cases, we can find out that the members from diamond community at the 2nd snapshot mainly inherits from diamond community at the 1st snapshot.

Conclusion:

In this paper, we introduced a new problem of analyzing the progression of community strengths. Community strength is a temporal measure which represents the probability that a particular community has a firm structure at the current snapshot. We proposed a two-stage framework which includes community detection and community strength analysis. This method can provide reliable and consistent community strength scores which are not only less sensitive to short-term noises in the current network but also adaptive to long-term networks evolution by considering the temporal smoothness. Moreover, the results of community strength analysis can help us find the top-K strongest or weakest communities throughout entire observation. Experiments on both synthetic and real dynamic datasets demonstrated the superior performance of the proposed method compared with other baseline methods in finding the top-K strongest or weakest communities.

References:

[1] M. Kolar, L. Song, A. Ahmed, and E. P. Xing, “Estimating timevarying

networks,” ArXiv e-prints, 2008.

[2] Y. Park and J. S. Bader, “How networks change with time,”

Bioinformatics, vol. 28, no. 12, pp. i40–i48, 2012.

[3] T. M. Przytycka, M. Singh, and D. K. Slonim, "Toward the dynamic interactome: it's about time," Briefings in Bioinformatics, vol. 11, no. 1, pp. 15–29, 2010.

[4] P. Bogdanov, M. Mongiovi, and A. Singh, "Mining heavy subgraphs in time-evolving networks," in Data Mining (ICDM), 2011 IEEE 11th International Conference on, 2011, pp. 81–90.

[5] M. Gupta, J. Gao, Y. Sun, and J. Han, "Integrating community matching and outlier detection for mining evolutionary community outliers," in In Prof. of KDD'12, 2012, pp. 859–867.

[6] Y. Chi, X. Song, D. Zhou, K. Hino, and B. L. Tseng, "On evolutionary spectral clustering," ACM Transactions on Knowledge Discovery from Data, vol. 3, no. 4, pp. 1–30, 2009.

[7] Y.-R. Lin, Y. Chi, S. Zhu, H. Sundaram, and B. L. Tseng, "Analyzing communities and their evolutions in dynamic social networks," ACM Transactions on Knowledge Discovery from Data, vol. 3, no. 2, pp. 1–31, 2009.

[8] A. L. Creekmore, W. T. Silkworth, and et al., "Changes in gene expression and cellular architecture in an ovarian cancer progression model," PLoS ONE, vol. 6, no. 3, pp. 1–16, 2011.

[9] N. Du, J. Gao, and A. Zhang, "Progression analysis of community strengths in dynamic networks," in Prof. of ICDM'13, 2013.

[10] L. J. Deborah, R. Baskaran, and A. Kannan, "A survey on internal validity measure for cluster validation," International Journal of Computer Science & Engineering Survey, vol. 1, no. 2, pp. 85–102, 2010.

[11] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," Physical Review E - Statistical, Nonlinear and Soft Matter Physics, vol. 69, no. 2 Pt 2, pp. 1–16,

2003.

[12] P. Paatero and U. Tapper, “Positive matrix factorization: A nonnegative factor model with optimal utilization of error estimates of data values,” Environmetrics, vol. 5, no. 2, pp. 111–126, 1994.

[13] C. Ding, T. Li, W. Peng, and H. Park, “Orthogonal nonnegative matrix t-factorizations for clustering,” In Prof. of KDD’06, pp. 126–135, 2006.

[14] P. Jaccard, “ ´ Etude comparative de la distribution florale dans une portion des Alpes et des Jura,” Bulletin del la Soci´et´e Vaudoise des Sciences Naturelles, vol. 37, pp. 547–579, 1901.

[15] L. Guan and M. Duckham, “Decentralized reasoning about gradual changes of topological relationships between continuously evolving regions,” in Spatial Information Theory, 2011, vol. 6899, pp. 126–147.

[16] S. Wu and X. Gu, “Gene network: Model, dynamics and simulation,” Computing and Combinatorics, vol. 3595, pp. 12–21, 2005.

[17] M.-S. Kim and J. Han, “A particle-and-density based evolutionary clustering method for dynamic networks,” In Proc. VLDB Endow., vol. 2, no. 1, pp. 622–633, 2009.

[18] S. E. Schaeffer, “Graph clustering,” Computer Science Review, vol. 1, no. 1, pp. 27–64, 2007.

[19] J. Leskovec, K. J. Lang, and M. W. Mahoney, “Empirical comparison of algorithms for network community detection,” In Prof. of WWW’10, 2010.

[20] A. Madan, M. Cebrian, S. Moturu, K. Farrahi, and A. S. Pentland, “Sensing the ’health state’ of a community,” IEEE Pervasive Computing, vol. 11, no. 4, pp. 36–45, 2012.

Polisetty A Satish Kumar, Student under M.Tech program II year C S E (Software Engg.) Newtons Institute of Engineering

G Ch Srinivasarao, Assistant

Assistant Professor, Dept of CSE, Newtons Institute of Engineering